

Low-Level-Fusion eines Laserscanner- und Stereokamera-Systems in der Fahrzeugumfelderfassung

Mathias Haberjahn

Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR)
Institut für Verkehrssystemtechnik
Rutherfordstraße 2, 12489 Berlin
eMail: mathias.haberjahn@dlr.de
URL: <http://www.dlr.de/ts>

Zusammenfassung. Ein Sensorsystem, bestehend aus einem 4-Ebenen-Laserscanner und einer Stereokamera, dient zur Erfassung des Fahrzeugumfeldes. Ziel ist es, Gefahrenpotentiale, zum Beispiel in Form weiterer Verkehrsteilnehmer, frühzeitig vorherzusehen. Im Folgenden wird ein neuartiger Ansatz zur konkurrierenden Fusion der Sensordaten auf der Punktebene vorgestellt. Innerhalb des Segmentierungsprozesses werden die Punktinformationen unter Einbeziehung der Punktgenauigkeiten und dem Segmentierungsverhalten der Sensoren fusioniert. Anschließend werden aus den so gewonnenen Segmenten Objektinformationen abgeleitet und der Bewegungsverlauf verfolgt. Der Genauigkeitsgewinn durch die Datenfusion im Vergleich zu den einzelnen Sensoren wird mit Hilfe einer GPS-basierten Referenzmessung dargestellt.

1 Einleitung

Der Einsatz von sensorgestützter Fahrerassistenz zur Erkennung und Vermeidung gefährlicher Situationen im Fahrzeugumfeld ist stark steigend. Der Grund hierfür liegt in der noch immer hohen Anzahl von schweren Verkehrsunfällen, ausgelöst durch menschliches Fehlversagen. Ziel muss es daher sein Systeme zu entwickeln, die in der Lage sind in kürzester Zeit ein Maximum an Gefahrenpotentialen zu erkennen und einzuschätzen und die Anzahl der Fehldetektionen auf ein Minimum zu reduzieren.

Beim alleinigen Einsatz typgleicher Sensoren lassen sich verfahrensbedingte Nachteile in der Umfelderfassung nur schwer kompensieren ([5], [8]). Es liegt also Nahe Sensoren mit unterschiedlicher Erfassungsart zu kombinieren. Hierbei lässt sich in kooperative und konkurrierende Sensor-Systeme unterscheiden. Die kooperative Sensorfusion beschreibt zumeist die unterstützende Rolle eines untergeordneten Sensors zu einem Hauptsensor ([1], [9], [12]). Gefahrenpotentiale die jedoch vom Hauptsensor nicht erkannt wurden, können somit auch vom Nebensensor nicht bewertet werden. Durch einen Ausfall des Hauptsensors wäre die gesamte Erfassung funktionsunfähig.

Durch die konkurrierende Sensorfusion versucht man diese Abhängigkeit zu vermeiden, indem alle Sensoren in der Lage sind, gleichrangige Informationen über die aktuelle Situation zu liefern [11]. Fehlende und falsche Detektionen sollen so

vermieden und die Güte der Situations-Beschreibung und Vorhersage durch die Sensorheterogenität gesteigert werden.

In dieser Arbeit wird die konkurrierende Sensorfusion zwischen einem 4-Ebenen-Laserscanner und einer Stereokamera auf der Rohdatenebene beschrieben. Der Aufbau und die Kalibrierung der Sensoren, sowie die Prozessierungsschritte von den Rohdaten zu den getrackten Objektinformationen werden in Abschnitt 2 beschrieben. Abschnitt 3 erläutert die eigentliche Datenfusion und die hierfür nötigen Vorschritte. Die Güte der Datenfusion im Vergleich zu den Einzelsensoren wird anhand von Referenzdaten im Abschnitt 4 dargestellt.

2 Aufbau der Sensoren

Für das Stereokamera-System werden zwei Monochrom-Kameras der Marke *PicSight GigE* von *Leutron Vision* mit einer maximalen Abtastrate von 20Hz bei einer Auflösung von 1392x1040 Pixel verwendet. Als Stereo-Matcher kommen vor allem der echtzeitfähige *Block Matcher* und der dichtere *Semi Global Block Matcher* von OpenCV zum Einsatz.

Der Laserscanner IBEO Lux besitzt 4 horizontale Scanebenen mit einem Abdeckungsbereich von 3.2° vertikal und 110° horizontal.

Zur Synchronisierung der beiden Sensoren sendet der Laserscanner in der Mitte des Scanvorganges einen Hardware-Impuls für eine umgehende Bildaufnahme an die beiden Kameras.

2.1 Kalibrierung der Sensoren

Das Stereosystem unterliegt einem dreistufigen Kalibrierungsprozess bestehend aus der Ermittlung der inneren Orientierung durch eine Bündelblockausgleichung, der anschließenden relativen Orientierung durch Bilddrehung und einer abschließenden absoluten Orientierung zur Maßstabsbestimmung und Ermittlung der Gesamtgüte der Kalibrierung [6].

Die Kalibrierung des Laserscanners und die Co-Registrierung zum Stereosystem werden in einem Schritt durchgeführt. Hierfür liefert das Stereo-System unter Laborbedingungen Referenzdaten in Form von Ebenen. Über ein Fehlermodell und eine räumliche Ähnlichkeitstransformation wird der Abstand zwischen den Laserpunkten und ihren Referenzebenen minimiert [6]. Durch die Co-Registrierung der Sensoren werden die Rohdaten des Laserscanner in das Bezugs-Koordinatensystem der linken Kamera transformiert.

2.2 Prozessierungskette

Die einzelnen Prozessierungsstufen zur Extraktion von getrackten Objektinformationen aus den Rohdaten sind mit Ausnahme der Kalibrierung und Filterung für beide Sensoren verfahrensgleich. Hintergrund hierfür ist der Aufbau einer Multi-Level-Fusion-Architektur zur dynamischen Kombination der Sensordaten auf unterschiedlichen Verarbeitungsniveaus (siehe Abb. 1).

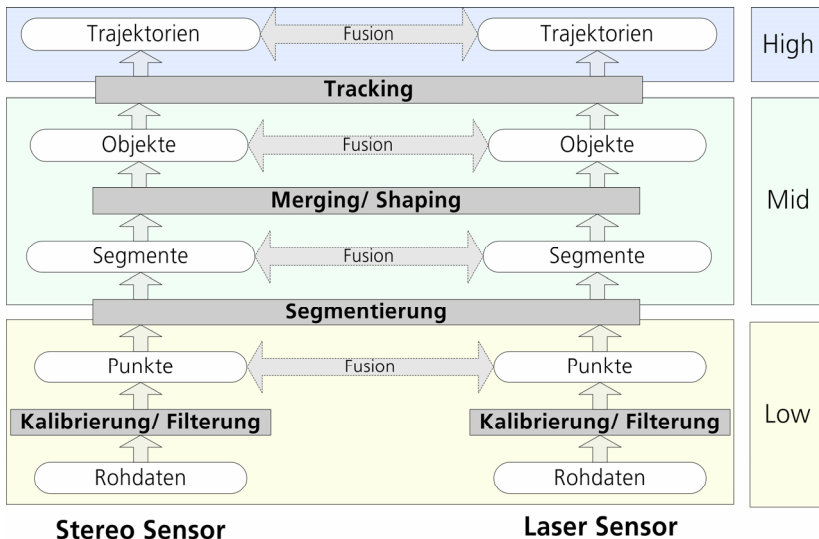


Abb. 1. Konkurrierende Prozesskette der Sensoren.

Die Filterung der Stereodaten beinhaltet mehrere Einzelschritte. Zum einen werden über eine Ebenenbestimmung der Fahrbahn alle sich darauf und darunter befindlichen Punkte entfernt. Zudem wird die gesamte Punktwolke reduziert, um die weitere Verarbeitung zu beschleunigen und fehlerhafte Punkte zu filtern (siehe Abschnitt 3.2). Analog dazu dient die Fahrbahnebene ebenfalls zur Filterung der Laserpunkte. Die Echobreite des empfangenen Laserimpulses dient zur Erkennung und Entfernung von Stördaten, z.B. verursacht durch Regen oder Nebel.

Durch die Segmentierung werden räumliche Punkthäufungen aus der Punktwolke als Segmente extrahiert (siehe Abschnitt 3.3).

Im Merging-Schritt werden die Segmente, die ein reales Objekt abbilden, zusammengefasst. Der Merging-Prozess ist eng an das Shaping gekoppelt. Hierbei wird für die

noch parallel zu den Koordinatenachsen ausgerichteten Objektboxen die räumliche Ausrichtung bestimmt. Über die Objektpunkte wird eine Ellipse in die Horizontalebene gelegt [4]. Die Ausrichtung der Ellipse liefert eine stabile Approximation für die Lage des Objektes und dient zur entsprechenden Neubestimmung der Objektbox [7].

Die so gewonnene Zusatzinformation über die Objektdrehung (ψ) geht neben den üblichen Positionsangaben des Objektmittelpunktes (x , y) in das Multi-Objekt-Tracking ein. Zusätzlich werden noch die Objektgeschwindigkeit (v) und Winkelgeschwindigkeit (ω) mitgeschätzt [2]. Die Hinzunahme des Drehwinkels und der Winkelgeschwindigkeit in den Zustandsvektor erlauben eine bessere Modellierung nicht-linearen Fahrverhaltens wie z.B. Kurvenfahrten. Das Tracking eines Objektes wird über ein nicht-lineares erweitertes Kalman Filter (EKF) realisiert. Als Ergebnis entsteht eine die Objektbewegung beschreibende Trajektorie für jedes verfolgte Objekt. Zusätzlich lässt sich über das Kalman-Filter ein Prädiktions-Anteil für jede Trajektorie bestimmen.

3 Fusion der Rohdaten

3.1 Genauigkeitsanalyse der Rohdaten

Zur optimalen Fusion der Punktinformationen der beiden Sensoren, sollte jedem Punkt eine Genauigkeitsangabe zu seiner Position zugeordnet sein. Den Ausgangspunkt hierfür bildet die Kalibrierung der Sensoren.

Wie im Abschnitt 2 erwähnt, werden zur Kalibrierung des Stereosystems drei aufeinander folgende Kalibrierschritte durchgeführt, die sich separat auf die Genauigkeiten der optischen Abbildung und der eigentlichen Punkt-Triangulation auswirken. Da die Parameter der inneren und relativen Orientierung in die Berechnung der absoluten Orientierung mit eingehen, lässt sich über diese eine Aussage über die Gesamtgenauigkeit der Kalibrierung ableiten. Als fehlerbehaftete Messgrößen werden die Objektkoordinaten herangezogen, die aus den Modellkoordinaten über die absolute Orientierung bestimmt wurden. Der Messfehler wird über die Rücktransformation dieses Objektpunktes und des zugehörigen Referenzpunktes in das Sensorkoordinatensystem der jeweiligen Kamera auf der Pixelebene messbar. Für eine ausreichend hohe Anzahl an Messpunkten lässt sich so das Streuverhalten für die Objektpunktbestimmung in den beiden Bilddimensionen ($\sigma_{x'}$, $\sigma_{y'}$) bestimmen. Über die Gleichungen zur Bestimmung eines Objektpunktes (X , Y , Z) aus den abbildenden Bildpunkten (x' , y') im Stereonormalfall kann man dieses Streuverhalten über Fehlerfortpflanzung in den Objektraum für die drei Dimensionen (σ_X , σ_Y , σ_Z) übertragen [10]:

$$\begin{aligned}\sigma_Z &= \frac{Z^2}{bc} \sigma_d, \quad \text{mit} \quad \sigma_d = \frac{\sigma_{x'}}{\sqrt{2}} \\ \sigma_X &= \sqrt{\left(\frac{x'}{c} \sigma_Z\right)^2 + \left(\frac{Z}{c} \sigma_{x'}\right)^2} \\ \sigma_Y &= \sqrt{\left(\frac{y'}{c} \sigma_Z\right)^2 + \left(\frac{Z}{c} \sigma_{y'}\right)^2}\end{aligned}\tag{1}$$

Somit lässt sich für jeden über das Stereo-Matching gewonnen Objektpunkt eine a-priori-Genauigkeitsangabe bestimmen.

Für die Laserpunkte werden die Genauigkeitsangaben ebenfalls aus der im Vorfeld durchgeführten Kalibrierung entnommen. Um die Varianzen der drei Koordinatenrichtungen zu ermitteln, werden die Abstände der kalibrierten Messpunkte zu ihren Schnittpunkten mit der Referenzebene betrachtet (Abb. 2).

Vereinfachend wird für beide Sensoren angenommen, dass das Streuverhalten der drei Koordinatenrichtungen untereinander nicht korreliert ist.

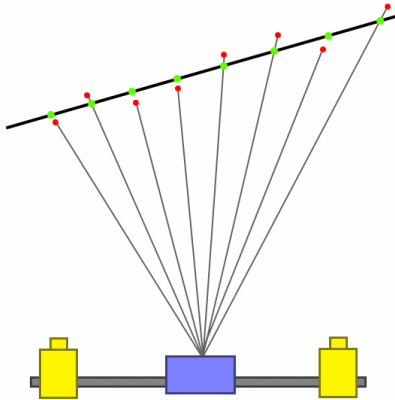


Abb. 2. Abweichung der Laserpunkte von zugehörigen Referenzpunkten auf der Referenzebene.

3.2 Reduktion der Stereodaten

Die Reduktion der Stereopunktwolke ist nötig, um die Informationsredundanz zu verringern und den anschließenden Segmentierungsprozess zu beschleunigen. Des Weiteren sollen hierbei grobe Ausreißer, verursacht durch fehlerhaftes Stereo-Matching, gefiltert und zusätzliche Angaben zur Genauigkeit der reduzierten Punkte erfasst werden.

Um die Reduktion der Punkt wolke möglichst gleichverteilt zu gestalten, wird das Disparitätenbild in quadratische Regionen gleicher Größe unterteilt. Innerhalb eines Fensters werden die Punktinformationen im Mittelpunkt kontrahiert. Die Streuung der Tiefeninformation im Fenster wird über ein Disparitätenhistogramm ermittelt. Ist hierbei die analytisch bestimmte Streuung (siehe Gleichung 1) für den Mittelpunkt geringer, so wird diese proportional angehoben. Somit werden Objektpunkte die Bereiche mit weniger Kontinuität in ihrer Tiefe abbilden und damit stärker Fehler im Stereo-Matching auslösen in ihrer Genauigkeit herabgesetzt.

Grobe Punktfehler und Randpunkte von Punkthäufungen, die die anschließende Segmentierung verzerren würden, sollen ebenfalls im Vorfeld der Segmentierung gefiltert werden. Hierbei wird eine Tiefenkarte in der Horizontalebene mit gleich großen Regionen erstellt. Um die Punktverteilungen möglichst zusammenhängend in der Tiefenkarte darzustellen, stellen die Zeilen der Tiefenkarte die möglichen Disparitätenstufen und die Spalten einen variablen metrischen Wert dar. Objektpunkte die in die gleichen Regionen fallen, werden akkumuliert. Durch einen distanzabhängigen Grenzwert wird die Tiefenkarte binarisiert, wodurch Regionen mit zu geringer Belegung als leer definiert werden. Objektpunkte die in leer definierten Regionen liegen, werden anschließend aus der Punkt wolke gefiltert (siehe Abb. 3).

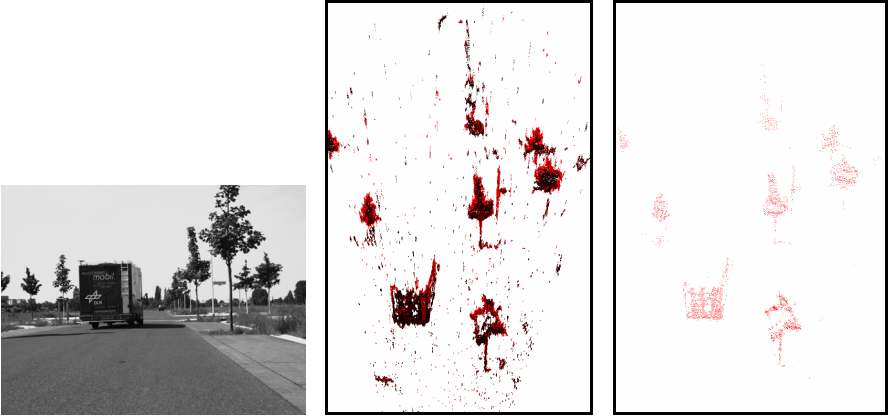


Abb. 3. Linkes Originalbild (Links). Stereo-Rohdaten (Mitte). Reduzierte Stereo-Punktwolke (Rechts).

3.3 Segmentierung der Punkte

Die reduzierten und bereinigten Punkte des Stereo Systems und die Laserpunkte werden im Zuge einer gemeinsamen Segmentierung von Punkte in Segmente fusioniert. Grundsätzlich unterscheidet sich die fusionierende Segmentierung nicht von dem Segmentierungsschritt für die beiden Sensorprozessketten. Nur die sensorabhängigen Angaben der Punktgenauigkeiten und Segmentierungs-Abständen müssen hierbei für beide Sensoren angegeben werden. Im Sinne der Sensorkonkurrenz ist somit gewährleistet, dass im Falle eines Sensorausfalls oder in nicht überlappenden Abdeckungsbereichen die Objektsegmentierung weiterhin funktionsfähig ist.

Die Segmentierung wird punktweise unter Berücksichtigung des Abstandes zweier Punkte vollzogen. Um den ersten Punkt wird dabei in Abhängigkeit vom Sensortyp und dem Abstand d zum Sensor ein Ellipsoid bestimmt, der den gültigen Segmentierungs-Bereich definiert (*S-Ellipsoid*). Die Halbachsenlängen a_s , b_s , c_s werden über einen konstanten o und distanzabhängigen Anteil sd definiert:

$$\begin{aligned} a_s &= (o_x + (s_x d))^2 \\ b_s &= (o_y + (s_y d))^2 \\ c_s &= (o_z + (s_z d))^2 \end{aligned} \quad (2)$$

Um die Genauigkeitsangabe der Punkte mit zu integrieren, wird nicht überprüft, ob ein Punkt im S-Ellipsoiden eines zweiten Punktes enthalten ist, sondern dessen Fehler-Ellipsoid (*F-Ellipsoid*) [3]. Das F-Ellipsoid gibt hierbei den Raum um einen Punkt an, in der sich die wahre Position des Punktes mit einer gewünschten sicheren Wahrscheinlichkeit p liegt (unter Annahme der Normalverteilung). Die Halbachsenlängen ergeben sich hierbei aus:

$$\begin{aligned}
 a_F &= \Phi^{-1} \left(\frac{p+1}{2} \right) \sigma_X + \mu_X \\
 b_F &= \Phi^{-1} \left(\frac{p+1}{2} \right) \sigma_Y + \mu_Y \\
 c_F &= \Phi^{-1} \left(\frac{p+1}{2} \right) \sigma_Z + \mu_Z, \quad \text{mit} \quad \mu_X \approx \mu_Y \approx \mu_Z \approx 0
 \end{aligned} \tag{3}$$

Zwei Punkte werden dann einem gemeinsamen Segment zugeschrieben, wenn für beide Punkte das F-Ellipsoid des einen Punktes im S-Ellipsoid des anderen Punktes enthalten ist. Somit ist es möglich innerhalb der Segmentierung sensorspezifisches Verhalten durch das S-Ellipsoid und die Punktgenauigkeit durch das F-Ellipsoid abzubilden.

Nach der Punktzuordnung an die Segmente, wird der minimal umschließende Quader mit paralleler Ausrichtung zu den Koordinatenachsen als Objektbox bestimmt (siehe Abb. 4).

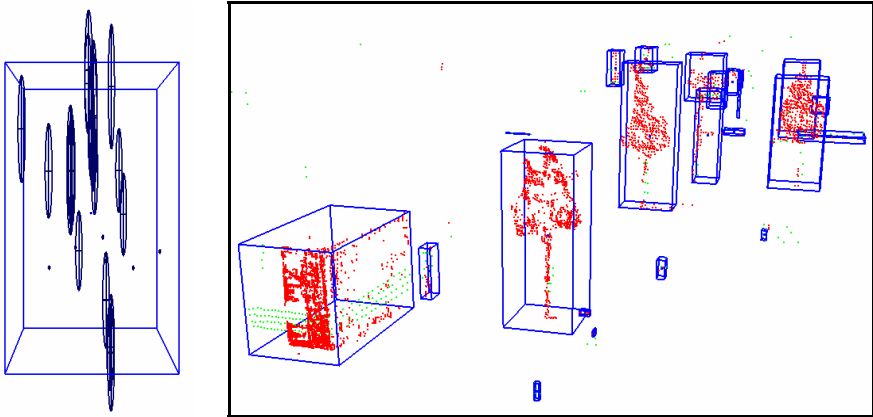


Abb. 4. F-Ellipsoide aus einer Segment-Draufsicht (Links). Gemeinsame Segmentierung der Stereopunkte (rot) und Laserpunkte (grün) (Rechts).

4 Ergebnisse

Um die Genauigkeit der Datenfusion unter realen Bedingungen zu testen, wurde ein zweites Fahrzeug als Referenzfahrzeug mit zwei hochgenauen GPS-Empfängern ausgestattet. Die beiden RTK-fähigen (*Real Time Kinematic*) GPS-Empfänger zeichnen ihre Position mit einer Frequenz von 10Hz und einer Streuung von unter 2cm auf. Die Position und Ausrichtung der beiden Antennen am Fahrzeug als auch das Fahrzeug wurden im Vorfeld genau vermessen. Somit ist es möglich unter der Annahme einer ebenen Fahrbahn die Fahrzeugform und somit auch die umschließende Objektbox aus den beiden Antennenpositionen abzuleiten.

Das eigentliche Messfahrzeug bleibt während der Messung stehen und zeichnet die Bewegungen des vor ihm fahrenden Referenz-Fahrzeugs auf.

Zum Vergleich der aus den Sensor-Daten und der Fusion abgeleiteten Objektboxen mit der Referenz-Objektbox, müssen die Referenzdaten in das Sensor-Koordinatensystem, welches durch die linke Kamera definiert ist, transformiert werden. Hierfür wird im Vorfeld der Messung die äußere Orientierung der linken Kamera im Bezug zum GPS-Koordinatensystem bestimmt. Für diesen Zweck wurden im Sichtbereich der Kamera verortete Messmarken platziert und über einen Rückwärtsschnitt ausgewertet.

Um das Verhalten der Sensor Fusion mit verschiedenen Fahrzeug-Entfernungen und Ausrichtungen zu testen, wurden eine Reihe von Fahrmanövern durchgeführt. Stellvertretend hierfür wird eine Fahrt mit Rechtsabbiegung ausgewertet (siehe Abb. 5).

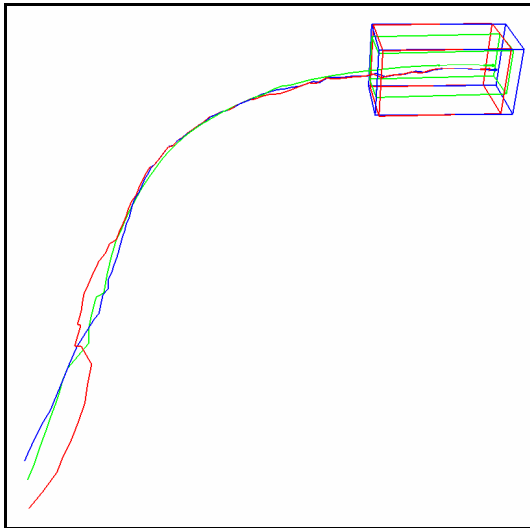


Abb. 5. Trajektorien des Stereo-Objektes (rot), des Laser-Objektes (grün) und des fusionierten Objektes (blau) vom abgebildeten Referenzfahrzeug.

Die fusionierten Objekte liegen im Schnitt mit ihren Objektmittelpunkten näher an am Mittelpunkt des Referenzobjektes als die entsprechenden Objekte der beiden einzelnen Sensoren (Abb. 6). Des Weiteren ist der Überlappungsgrad von Referenzobjekten und Fusions-Objekten höher verglichen mit den Sensor-Objekten (Abb. 7). Der Überlappungsgrad berechnet sich hierbei aus dem Produkt der Verhältnisse von Überlappungsvolumen zu Referenzboxvolumen und Überlappungsvolumen zu Sensorboxvolumen und beträgt bei Deckungsgleichheit 1 und bei getrennten Objektboxen 0. Das hierbei vergleichsweise schlechte Abschneiden der Laserobjekte erklärt sich aus der ungenauen Objekthöhenmessung aufgrund der geringen Anzahl der Scanebenen .

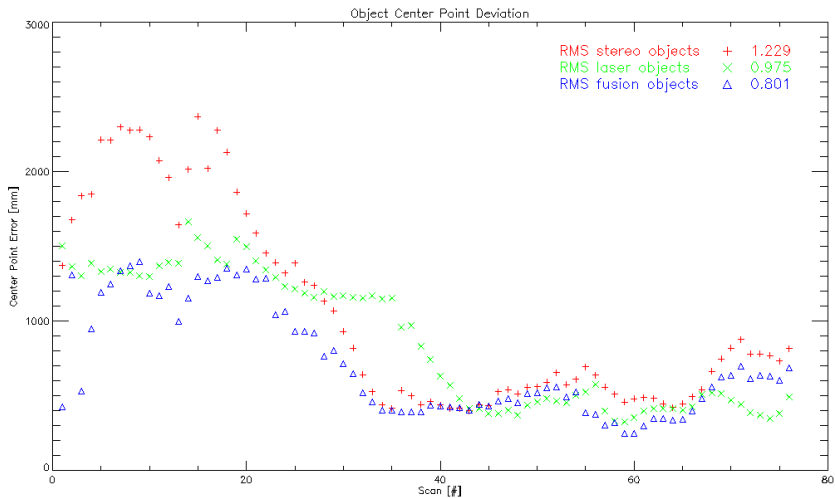


Abb. 6. Abweichung der Objektmittelpunkte von Referenzobjekten zu Sensor- und Fusions-Objekten.

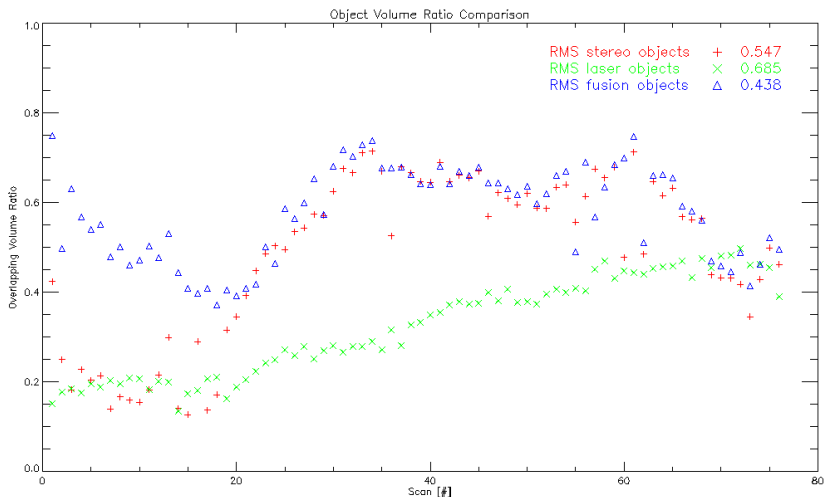


Abb. 7. Überlappungsgrad von Referenzobjekten und Sensor- und Fusions-Objekten.

5 Zusammenfassung und Ausblick

Es wurde ein neuartiges Verfahren zur Fusion der Rohdaten eines 4-Ebenen-Laserscanners und eines Stereokamerasystems vorgestellt. Neben der räumlichen Verteilung der Punkte steuern die Genauigkeitsangaben der einzelnen Punkte und die Modellierung des Segmentierungsverhaltens der beiden Sensoren die Fusion.

Die Ergebnisse haben gezeigt, dass durch die Fusion die Genauigkeit der Objekterkennung im Vergleich zu den Einzelsensordaten erhöht wurde.

Zusätzlich kann davon ausgegangen werden, dass durch die Sensorkonkurrenz die Ausfallwahrscheinlichkeit des Systems verringert wurde. Des Weiteren kann man annehmen, dass ebenso die Anzahl der Fehl- und Falsch-Detektionen durch die höhere Stabilität der Objekthypothesen verringert wurden, was jedoch noch durch entsprechende Versuche belegt werden muss.

In Zukunft wird die Daten-Fusion auf Objekt- und Trajektorien-Ebene vorangetrieben, um darauf ein dynamisches Multi-Level-Fusion-System aufzubauen.

Literatur

1. Allesandretti, G., Broggi, A., Cerri, P.: Vehicle and Guard Rail Detection Using Radar and Vision Data Fusion. IEEE Trans. Intell. Transp. Syst., vol. 8, pp. 95 (2007).
2. Boehringer, F.: Gleisselektive Ortung von Schienenfahrzeugen mit bordautonomer Sensorik. Universität Karlsruhe, PhD: 178 (2008).
3. Eberly, D.: Intersection of Ellipsoids. Geometric Tools (2000).
4. Fitzgibbon, A., Pilu, M., Fisher, R.B.: Direct Least Square Fitting of Ellipses. IEEE Transactions of Pattern Analysis and Machine Intelligence, v.21 n.5, 476-480 (1999).
5. Fuerstenberg, K.Ch., Dietmayer, K.: Pedestrian Recognition and Tracking of Vehicles using a vehicle based Multilayer Laserscanner. 10th World Congress on Intelligent Transport Systems, Madrid, Spain (2003).
6. Haberjahn, M., Reulke, R.: Cross-Kalibrierung eines Mehrzeilen-Laserscanner- und Stereokamera-Systems zur Fahrzeugumfelderfassung. 3D-Nordost, Berlin, Germany (2009).
7. Haberjahn, M., Reulke, R.: Vehicle Object Discrimination and Tracking by a combined Laser Scanner Stereo System. The 10th Asian Conference on Computer Vision, Vol. 6468-6469 (2010).
8. Huang, Y., Fu, S., Thompson, C.: Stereovision-based object segmentation for automotive applications. In EURASIP Journal on Applied Signal Processing, pages 2322-2329 (2005).
9. Kaempchen, N., Dietmayer, K.: Fusion of laserscanner and video for advanced driver assistance systems. In 11th World Congress on Intelligent Transportation Systems, Nagoya (2004).
10. Kraus, K.: Photogrammetrie, Geometrische Informationen aus Photographien und Laserscanneraufnahmen. Band 1, 7. Auflage, Walter de Gruyter (2004).
11. Labayrade, R., Royere, C., Gruyer, D., Aubert, D.: Cooperative Fusion for Multi-Obstacles Detection With Use of Stereovision and Laser Scanner. Proc. Int. Conf. Adv. Robot., pp. 1538 (2003).
12. Perollaz, M., Labayrade, R., Royere, C., Hautiere, N., Aubert, D.: Long Range Obstacles Detection Using Laser Scanner and Stereovision. Procs. IEEE Intelligent Vehicles Symposium 2006. Tokyo, Japan, 182-187 (2006).